Mapping the properties of bidentate ligands with calculated descriptors (LKB-bid)

Natalie Fey,*[‡] Alexander Koumi,[‡] Andrei V. Malkov,[§] Jonathan D. Moseley,[$] Bao N. Nguyen,[v] Simon N. G. Tyler,[$] Charlotte E. Willans[v]

[‡]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK; [§] Department of Chemistry, Loughborough University, Epinal Way, Loughborough, Leicestershire, LE11 3TU, UK; [$] CatSci Ltd., CBTC2, Capital Business Park, Wentloog, Cardiff, CF3 2PX, UK; [v]School of Chemistry, University of Leeds, Leeds, LS2 9JT, UK.

Abstract

We have extended the Ligand Knowledge Base (LKB) approach to consider a broad range of bidentate ligands, varying donors, substituents and backbones, which gives rise to a diverse set of 224 ligands in a new database, LKB-bid. Using a subset of steric and electronic parameters described previously for bidentate P,P-donor ligands (LKB-PP), here this approach has been applied to a wider set of bidentate ligands, to explore how these modifications affect the properties of organometallic complexes. The resulting database has been processed with Principal Component Analysis (PCA), generating a "map" of ligand space which highlights the contribution of donor atoms and bridge length to the variation in ligand properties. This mapping of bidentate ligand space with DFT-calculated steric and electronic parameters has demonstrated that the properties of ligands with different donor atoms can be captured within a single computational approach, providing both an overview of ligand space and scope for the more detailed investigation and comparison of different ligand classes.

Introduction

While some homogeneous organometallic catalysts have found widespread use, for example in cross-coupling, metathesis, hydroformylation and hydrogenation, one of the key challenges in this area remains the selection of suitable ligands for a given synthetic task.[1, 2] Ligand selection can aim to optimise catalyst properties, most commonly activity, selectivity and stability, although other figures-of-merit for the catalyst, such as cost, toxicity and solubility, may also be of interest.

Some of the most successful examples of homogeneous catalysis have been found to be viable with different ligand classes, as shown, for example, by the evolution of ruthenium metathesis catalysts from P-donor spectator ligands to N-heterocyclic carbenes, with the latter also supported by chelating benzylidene ethers and other types of ligand.[3] Similarly, in copper-catalysed Ullmann cross-coupling reactions the choice of ligand can be used to modify selectivity in aminoalcohol substrates; in this case, N,N- and O,O-donor ligands can be used to control selectivity, although the mechanistic considerations are complex.[4] Detailed mechanistic studies of catalytic reactions can shed light on why different catalysts alter activity and selectivity and help to elucidate ligand effects.[5] However, such studies are time-consuming and thus generally remain focussed on comparing a small number of systems in considerable detail, rather than large-scale evaluation and prediction of ligand effects on catalysis,[6] so alternative approaches, avoiding full mechanistic studies, remain desirable.[7]

To guide catalyst optimisations, ligand steric and electronic parameters have been reported by a number of authors for different ligand classes, using both experimental and calculated data, and their

utility for guiding catalyst optimisation has been described in a recent review.[2] When calculated, ligand steric and electronic parameters can also be used to set novel ligand designs into context,[8] and thus suggest possible applications.[6, 9] Tolman's early work on P-donor ligands[10] also considered extension of his cone angle steric parameter to bidentate ligands, and other attempts to consider several types of ligands in a single data framework have been reported (recently reviewed in [2]). However, most of these parameters are specific to a single type of ligand, often with limited transferability. Early comparisons between P-donors and carbenes hinted at the problems with developing an approach suitable for their comparison.[11, 12] For electronic properties in particular, a range of metal complexes had to be considered for the determination of CO stretching data (note that related experimental work has been reviewed recently in reference [13]), eventually leading to the identification of a transferable calculated electronic parameter based on a range of iridium complexes.[11] However, such combined approaches remain rare and most ligand classes are considered separately/in isolation.

A more general, quantitative tool for experimental design, guiding catalyst discovery and optimisation to promising candidates across a wide range of chemical space as accessed by changing ligands, would be highly desirable. In such an approach, parameters should be relatively easy to determine, ideally by computation to allow the consideration of novel designs before synthesis, as well as transferable to different coordination environments and ligand donor atoms. Here, we have extended the Ligand Knowledge Base (LKB) approach, developed by a consortium of research groups in Bristol,[14, 15] to consider a broad range of bidentate ligands, varying donors, substituents and backbones. This gives rise to a diverse set of 224 ligands. Using a subset of steric and electronic parameters described previously for bidentate P,P-donor ligands (LKB-PP),[16-18] this approach has been applied to explore how these modifications affect the properties of organometallic complexes. Insights from this exploration of wider bidentate chemical space and some potential applications for this database (LKB-bid) in experimental design and catalyst optimisation are also discussed.

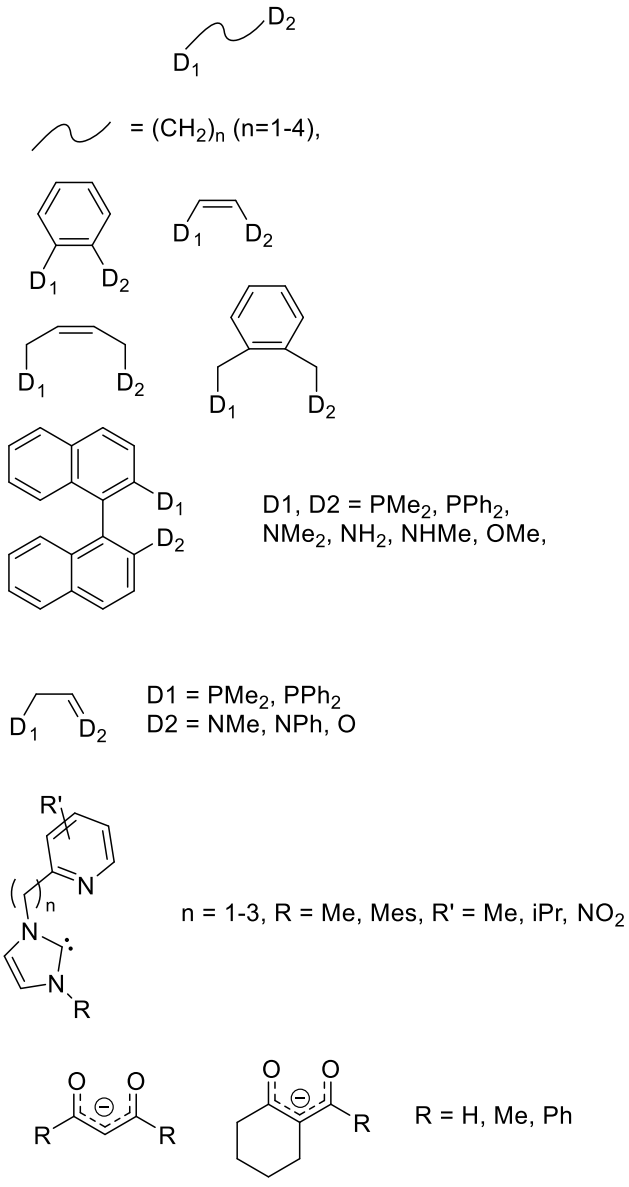
Database Design

The LKB approach has been described extensively elsewhere, including the development of a database for bidentate P,P- and P,N-donor ligands (LKB-PP)[16, 17] and its application to screening backbone and substituent effects (LKB-PP_{screen}).[18]

In brief, this approach uses a standard DFT approach (see ESI for full computational details and sample files) to optimise the free ligand,[14] as well as a range of representative complexes; calculations at this level of theory can capture relatively subtle electronic effects, important to catalyst properties. From these ground state geometry optimisations, data are harvested. Some of these parameters (note that the term descriptors will be used interchangeably throughout) deliberately capture isolated steric and electronic effects, while others measure combined effects on complex properties. The resulting database, LKB-bid in the present case, is then analysed with a statistical data projection approach, principal component analysis (PCA),[19] to reduce the dimensionality and facilitate visualisation. This is possible because the descriptors are reasonably highly correlated, as they are derived from the same ligand in different environments, and PCA can be used to derive a new set of uncorrelated variables (principal components, PCs), which are linear combinations of the original descriptors and capture most of the variation in the dataset in fewer dimensions. Plots of the first few PCs allow the generation of so-called maps of chemical space,[20] and both individual descriptors and PCs can also be used in

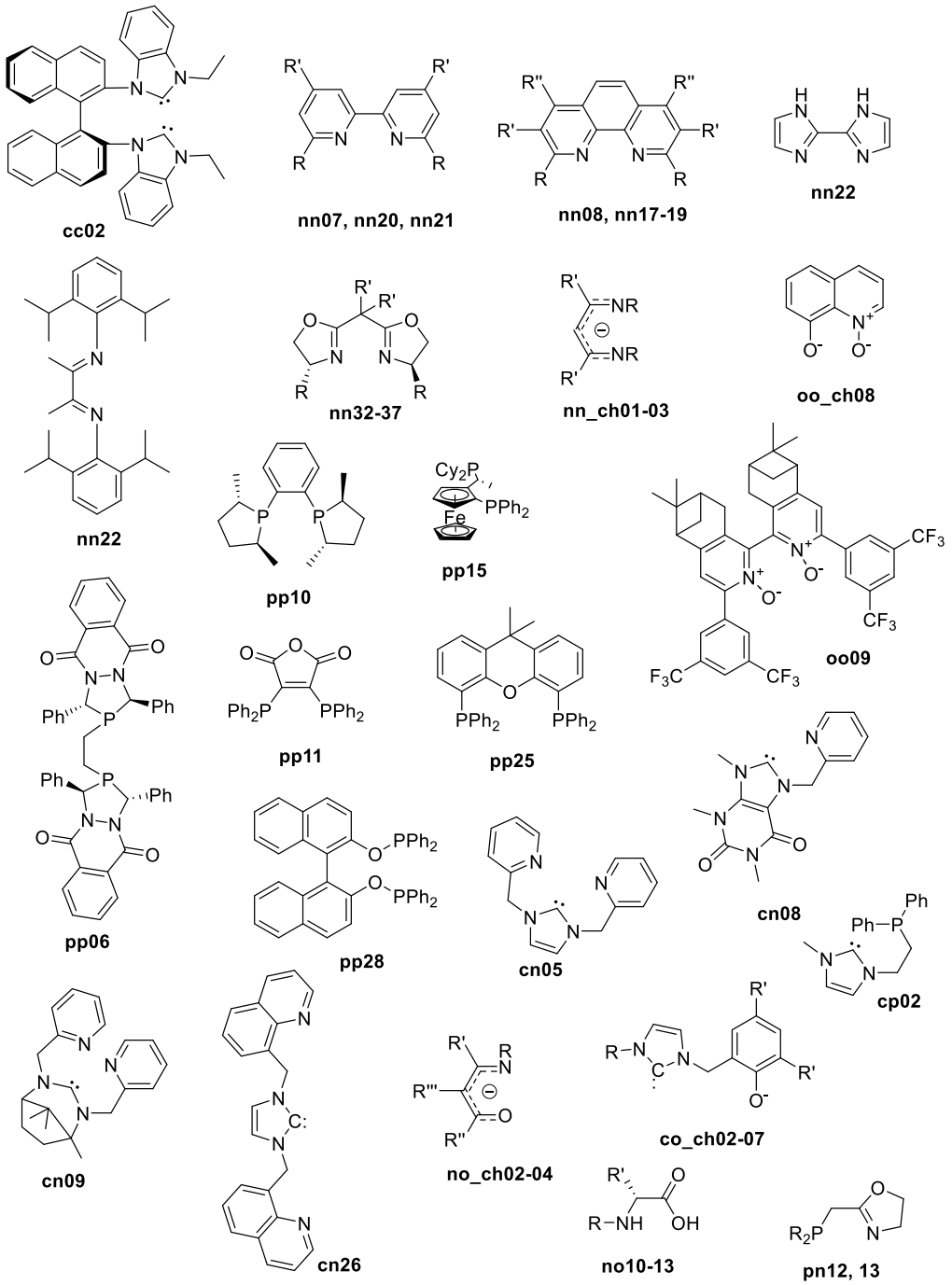
multivariate regression models where a suitable figure-of-merit/response can be identified, allowing both the interpretation of data, and, at times, prediction for related or novel systems.[15, 17, 21, 22]

In common with earlier work,[14] a standard DFT approach (BP86/6-31G*, with LACV3P on metal atoms; for full references and workflow, see computational details in ESI) has been used. The functional used overbinds slightly, compensating for the lack of dispersion corrections and ensuring that most geometry optimisations are successful. We have previously shown that the steric parameters used can capture some of the differences between standard and dispersion-corrected data.[23] Descriptors are based on potential energies, avoiding the more expensive frequency calculations. The chelate effect[24] and ligand hemilability can be challenging to capture computationally, even with more sophisticated considerations of solvent and dispersion effects, along with thermochemical corrections to capture entropy, and such corrections have not been attempted on the scale of this database. It is worth bearing in mind, however, that trends in ligand effects are likely to be reasonably robust, even where energy effects are not fully captured.



Scheme 1: Main ligand variations explored. See ESI for full list of ligands and labels used (Table S1).

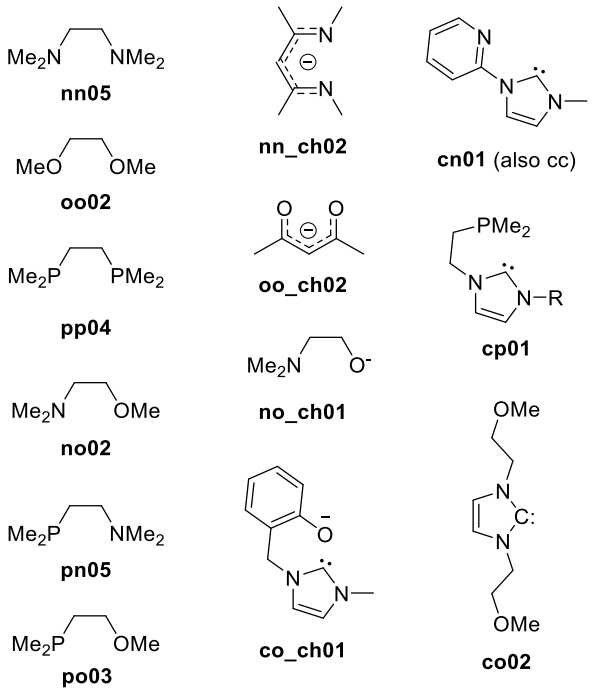
PCA is not statistically robust,[21] making analyses sensitive to outlier observations and ligands were chosen to be representative of different chemistries, seeking to balance different donor combinations. Systematic variations of donors, substituents and backbones (Scheme 1) have been combined with chemically interesting ligands drawn from published work (Scheme 1), balancing synthetic relevance and chemical adventure. Where several coordination modes and protonation states could be considered for the ligand, we have used the most likely/stable one. The full ligand list and labels have been included in the ESI (Table S1), while Schemes 1 and 2 provide an overview of the main structural variations considered. Some charged ligands have been included, as have representative monodentate ligands to capture the effect of the backbone on ligand properties.



Scheme 2: Examples of additional ligands included in LKB-bid. See ESI for full list and detailed labels.

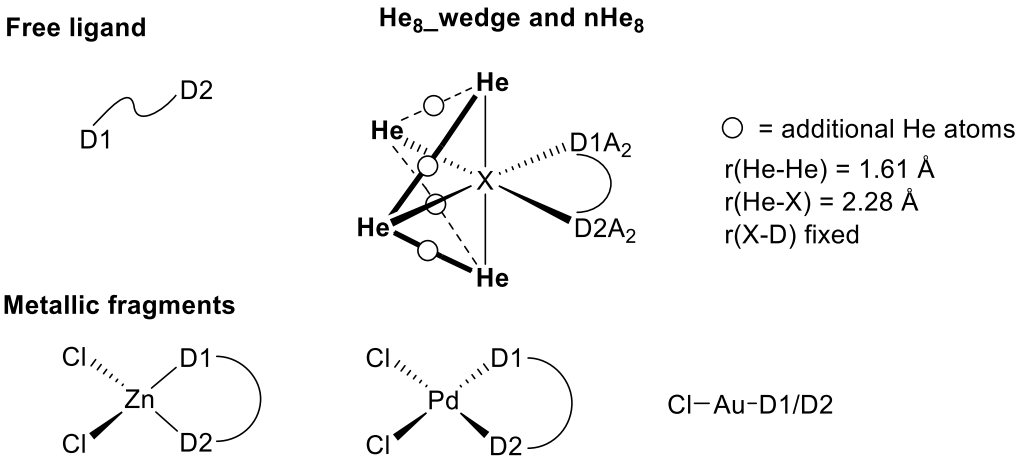
With a view to utilising ligand data already reported, the current database relies on some of the descriptors selected for LKB-PP,[17] which have a bias towards late transition metal complexes, discussed further below. However, the "split ligand descriptors" developed previously,[16, 17] capturing the proton affinity, HOMO and LUMO energies for each individual donor atom with suitable truncation, became increasingly difficult to construct as ligand variation increased; these descriptors have thus been dropped from consideration here. They correlate quite highly with other descriptors capturing σ-donicity, e.g. those derived from the [$ZnCl_2$] fragment, such that these properties continue to be captured. The distances between donor atoms and the centre of the He_8 wedges used to calculated steric parameters have been adjusted to take account of different donor atom types ("D–X" distance = 2.28 Å (P), 2.00 Å (C, N), 2.05 Å (O)). As described previously,[16, 17] these steric parameters mimic the steric hindrance of an idealised octahedral coordination environment by placing a wedge of helium atoms where *cis* ligands would be found. This allows the calculation of a computationally-convenient, repulsive interaction energy for ligands in their chelating conformation. The He_8_wedge parameter freezes the donor atom positions as optimised in the [$ZnCl_2$] complexes, found to provide a reasonable approximation of the "natural" bite angle, while the nHe_8 calculation allows for movement of donors in response to steric pressure, albeit constrained to the D-X distances noted, designed to capture whether ligands can respond to steric hindrance.[17] In addition, we have introduced the binding energy of the [AuCl] fragment,[15, 25] calculated for individual donor atoms, as a possible descriptor for ligand hemilability to this database as well.

We note that in LKB-PP, the P,N-donor ligands considered appear slightly separate from P,P donors in the chemical space sampled by LKB-PP, suggesting that the difference in donor atoms, dampened by using the change in metal-donor distance compared to a reference ligand, is nonetheless an important source of variation in the data which has been captured by the PCA analysis.[15, 16] In the present database, consideration of a range of different donors, including carbenes and phosphines, highlighted that the main difference arises from the strength of metal-ligand binding. This can be useful in some contexts, but in the present case the focus should be on the effect of ligands on the properties of the metal complexes, i.e. treating them as spectator ligands. Rather than using measured metal-donor distances, the change compared to a representative reference ligand (Scheme 3 and highlighted in green in ligand list, see ESI) for each donor set has been used to reduce the impact of different binding modes on the data analysis. Where chemically relevant, these references have two atoms in the backbone between donors and methyl substituents; in some cases (cp01, co02, nn_ch02, oo_ch02, co_ch01), the nearest comparable ligand was used instead.



Scheme 3: Reference ligands used.

Scheme 4 shows the complexes calculated and Table 1 summarises the data harvested from these calculations. The complete database is available as part of the ESI.



Scheme 4: Complexes calculated for LKB-bid.

Table 1: Descriptor data harvested from calculations on complexes (Scheme 4).

Descriptor	**Derivation (Unit)**
Free Ligand	
He_{8}_wedge	Interaction energy between ligand in chelating conformation and wedge of 8 He atoms, maintaining donor atom position similar to [$ZnCl_2$] complex,[a] E_{He8_wedge} = $E(He_8(D1{\sim}D2)) - E(He_8) - E((D1{\sim}D2))$ (kcal mol^{-1})

nHe_8	Interaction energy between ligand in chelating conformation and wedge of 8 He atoms, donor atoms free to move at fixed X-D distances,[b] $E_{n\mathrm{He8}}$ = $E(He_8(D1{\sim}D2)) - E(He_8) - E((D1{\sim}D2))$ (kcal mol^{-1})
Zinc complexes Zn(D1∼D2)Cl_2	
BE(Zn)	Bond energy for dissociation of D1∼D2 ligand from metal fragment (kcal mol^{-1}) [c]
Zn–Cl	Average Zn–Cl distance (Å)
∠D1–Zn–D2	Ligand bite angle in complex (degrees)
ΔD1–R(Zn), ΔD2-R(Zn)[c]	Change in average D–R distances cf. free ligand (Å)
ΔZn–D1, ΔZn–D2	Change in Zn–D distances cf. reference ligand (Scheme 4) (Å)
Q(Zn)	NBO charge on $ZnCl_2$ fragment
Palladium complexes Pd(D1∼D2)Cl_2	
BE(Pd)	Bond energy for dissociation of D1∼D2 ligand from metal fragment (kcal mol^{-1}) [d]
Pd–Cl	Average Pd–Cl distance (Å)
∠D1–Pd–D2	Ligand bite angle in complex (degrees)
ΔD1–R(Pd), ΔD2-R(Pd)[c]	Change in average D–R distances cf. free ligand (Å)
ΔPd–D1, ΔPd–D2	Change in Pd–P distances cf. reference ligand (Scheme 4) (Å)
Q(Pd)	NBO charge on $PdCl_2$ fragment
Gold Complexes ([AuClL], not included in main PCA)	
BE(Au, D1), BE(Au, D2)	Bond energy for dissociation of single donor from [AuCl] fragment (kcal mol^{-1})[d]

[a] Donor atoms in fixed positions, fixed "D–X" distance = 2.28 Å (P), 2.00 Å (C, N), 2.05 Å (O) ; [b] Fixed "D–X" distances (as for He_8_wedge[a]), D atom position free to move; [c] R = Substituents on D atoms, [d] BE = $[E_{tot}(\text{fragment})+E_{tot}(L)]-E_{tot}(\text{complex})$

Some basic descriptive statistics for key descriptors are shown in Table 2 for the larger subsets (N>10) of ligands with different donors. These confirm that the chosen descriptors are responsive to chemical changes yet avoid excessive outliers. The full analysis can be found in the ESI (see spreadsheet). In addition, we have considered bivariate correlations between descriptors (see spreadsheet). These are largely as might be expected, *i.e.* high for descriptors capturing similar behaviours in different complexes (BE, bite angles) and for those related through electronic effects (*e.g.* between BE and ligand *trans* influences). Changes in M-L distances compared to reference compounds are not showing high linear correlations (>0.5) with other descriptors, but especially those derived from the Pd complex

show some correlation with descriptors related to steric and σ-donation effects, again as expected from our understanding of metal-ligand interactions.

Table 2: Mean (standard deviation) for representative descriptors by donors (only showing those with N>10, see spreadsheet for full table).

D1,D2	He_8_wedge	nHe_8	BE(Zn)	BE(Pd)	∠D1–Zn–D2
C,N (N=38)	21.0 (7.4)	19.1 (5.7)	54.3 (5.2)	94.3 (6.7)	91.9 (11.1)
N,N (N=41)	14.5 (7.8)	13.8 (7.1)	46.4 (5.6)	73.3 (7.6)	81.3 (9.2)
N,O (N=20)	10.1 (4.8)	8.5 (5.3)	36.2 (5.1)	55.2 (6.2)	79.7 (9.4)
O,O (N=11)	6.1 (2.5)	6.4 (3.6)	34.8 (6.3)	44.3 (6.4)	80.6 (8.9)
P,P (N=25)	22.5 (9.3)	18.3 (5.5)	33.7 (5.0)	89.5 (8.3)	88.9 (12.1)
P,N (N=30)	16.3 (7.0)	14.0 (5.7)	37.5 (3.8)	79.0 (6.6)	82.7 (11.4)
P,O (N=16)	10.4 (4.0)	9.2 (3.5)	32.7 (3.2)	65.5 (4.1)	79.8 (8.9)

Maps of Bidentate Ligand Space

LKB-bid can be processed further, focussing on the use of PCA to aid visualisation of the ligand space mapped. Here we first consider all ligands (section a), before focussing on a subset of ligands with C, N and O donors (section b).

a. All Ligands

Eighteen ligand descriptors have thus been calculated as described above for 217 bidentate ligands as well as 7 cases combining two monodentate ligands, assuming a *cis*-, bisligation in all complexes. In addition, binding energies for the coordination of the [AuCl] fragment to individual donor atoms were calculated for most ligands. For some C,N donor ligands (cn01-03, 06), the gold fragment migrated from N to C, so a binding energy could not be calculated. These descriptors have thus been left out of the main analysis but will be considered further below.

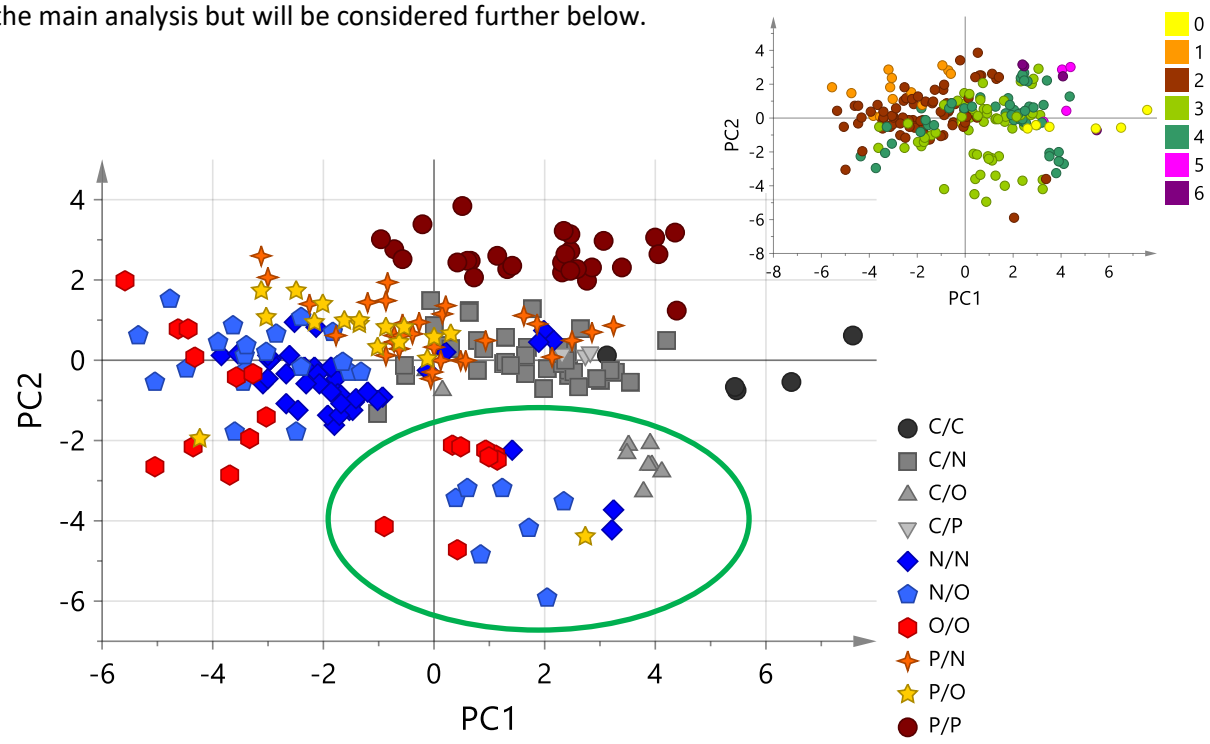


Figure 1: PC score plot, showing PCs 1 and 2, for the full ligand set (224 ligands). Each marker corresponds to a ligand and these first two PCs capture 52% of the variation in the database. Colour-coding according to donor atoms for main plot, and according to backbone length for inset. Green ellipse marks charged ligands. See ESI for larger version of main plot, showing all ligand numbers.

Figure 1 shows the ligand scores for the database based on a set of eighteen ligand descriptors, i.e. excluding those derived from the [AuCl] fragment. This was obtained by PCA of the correlation matrix, and Figure 1 displays the first two PCs, which capture 52% of the variation in the database.

Table 3 shows descriptor loadings for PCs 1-3, illustrating that most descriptors load highly onto the first two PCs (see also Fig. S1 for a plot of descriptor loadings; see ESI for all diagnostics (Table S2, Figure S1, spreadsheet) and detailed plots (Figures S2-S8)). As noted earlier, PCA is not statistically robust, with different ligand sets altering the composition of each PC, which are linear combinations of descriptors used. We note that bite angles, M-Cl distances and the He_8 steric descriptors load more highly on PC1, suggesting perhaps that steric and σ-bonding effects are more strongly captured in this dimension. However, most descriptors load on both PC1 and PC2 (Table 3 and ESI), making a more detailed interpretation challenging in this case.

Table 3: Descriptor loadings for PCs 1-3.

	PC1	**PC2**	**PC3**
% variation captured	**35.9**	**15.6**	**12.6**
He_8_wedge	0.296	0.184	-0.307
nHe_8	0.286	0.145	-0.318
BE(Zn)	0.212	-0.451	0.106
Zn–Cl	0.332	-0.261	-0.022
ΔZn–D1	-0.020	0.190	0.128
ΔZn–D2	-0.024	0.092	0.029
ΔD1–R(Zn)	-0.219	-0.220	-0.374
ΔD2–R(Zn)	-0.117	-0.355	-0.183
∠D1–Zn–D2	0.313	-0.068	-0.313
Q(Zn)	-0.256	0.271	-0.196
BE(Pd)	0.316	-0.144	0.297
Pd-Cl	0.343	0.066	0.163
ΔPd–D1	0.163	0.306	-0.170
ΔPd–D2	0.185	0.148	-0.248
ΔD1–R(Pd)	-0.206	-0.242	-0.376
ΔD2–R(Pd)	-0.120	-0.305	-0.196
∠D1–Pd–D2	0.292	0.007	-0.270
Q(Pd)	-0.170	0.289	-0.091

Data points close to each other on this map indicate greater similarity, as measured by the descriptors used, while those further apart are more different. Note that 18 descriptors will give rise to 18 PCs, albeit of decreasing importance, and additional dimensions beyond PCs 1 and 2 may be important here. This has been discussed further in the ESI (Figures S9-S11); here we will focus on the first two PCs to simplify visualisation and discussion.

The PC score plot shows a separation of data according to donor atoms, with P,P-donor ligands most distant from other ligand types in the top right hand corner/North East of the map (in line with calling these plots "maps", a geographical naming convention is useful here). With PCA designed to identify the largest sources of variation in the data, the differentiation of the second row, larger P-donor atoms is not surprising and aligns well with our chemical knowledge. Most C,C-donor ligands have moved away from other systems as well, along the equator (PC1>2, PC2≈0), presumably driven by higher M-C binding energies than the other ligands considered, but only a few examples have been included here and such undersampling is hampering more detailed conclusions from being drawn.

Compared to ligands with the same donors, mixed donor systems have more similar properties across different donor atoms, shown by an overlap of data points from different ligand classes near to the equator and towards the South West; this is particularly pronounced for N,N-, N,O- and O,O-ligands, but P,N- and C,N-donor ligands also show some overlap, at least in terms of these first two principal components. (Note that this is true when considering PC3, as well, Figures S9 and S11.) Finally, there is a group of ligands slightly separate from others in the South East quadrant of the map (PC1>0, PC2<0, marked with green ellipse in Fig. 1), close to the central meridian. These ligands are charged, giving them higher ligand binding energies to metal fragments and illustrating how the PCA approach is sensitive to the main variations in data. (The ESI also includes PCA maps excluding these ligands, Tables S3, Figures S12, S13.)

South West to North East banding of ligand types arises from the variation in donors, while, as shown as an inset in Figure 1, variation from North West to South East, i.e. across the other diagonal of this map, is to some extent related to backbone length. This becomes even more obvious when excluding charged ligands (see ESI, Figures S11, S12); a similar spread of data according to backbone was observed for LKB-PP_{screen}.[18] Note, however, that the sampling of longer backbone lengths is uneven across different ligand classes. In addition, bridge flexibility, facilitating the adoption of a chair-like backbone conformation, as well as length, can influence the ligand position, as shown in Figure S8b, which suggests that more flexible ligands and longer bridges are likely to be found towards the Eastern side of the map for each ligand class.

b. C-, N- and O-Donor Ligands

As noted in a recent review,[2] perhaps the computationally most well-sampled ligand classes in catalysis are monodentate P-donor and carbene ligands, with bidentate P,P-donor ligands a relatively distant third. That notwithstanding, other ligand classes play important roles in this field, especially for supporting earth abundant, first row transition metal centres, where carbenes along with N- and O-donor ligands tend to be popular.[26]

With a view to exploring these subsets of ligands in greater detail, the PCA was repeated after removal of ligands containing at least one P-donor. The map resulting from consideration of the first two PCs is shown as Figure 2 and now captures 57 % of the variation in the database for 146 ligands. (See ESI

for details and diagnostics) With the P-donor ligands, which lie away from other bidentate ligands (Figure 1) due to familiar differences in M-L bonding, removed from consideration, resolution of the remaining ligand clusters is improved, facilitating more detailed analysis of similarities and differences for the remaining ligands.

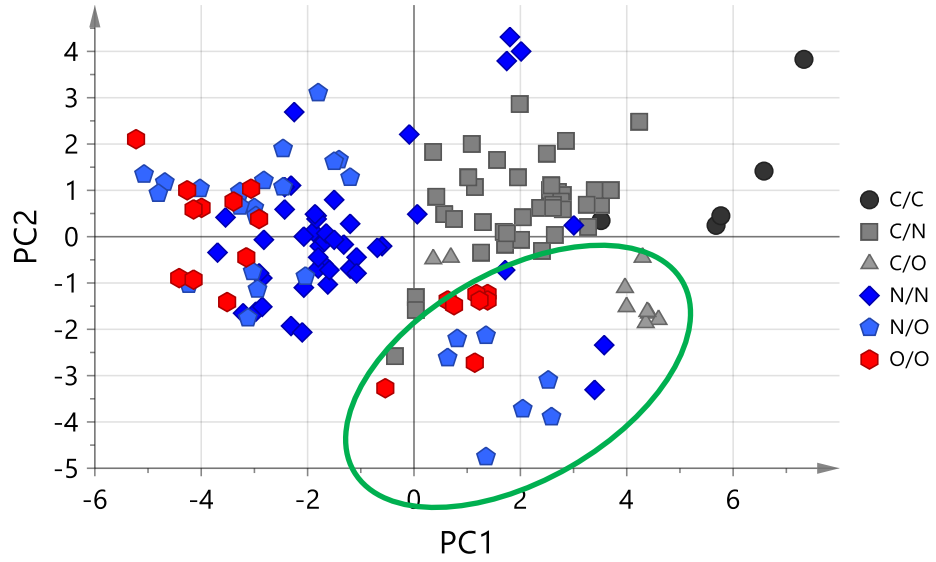


Figure 2: PC score plot, showing PCs 1 and 2, for ligands with C-, N- and O-donors only (146 ligands). Each marker corresponds to a ligand and these first two PCs capture 57% of the variation in the database. Colour-coding according to donor atoms. Green ellipse marks charged ligands. See ESI for larger version of main plot, showing all ligand numbers.

Again, clustering according to donors (Figure 2) and banding according to backbone length and flexibility (ESI Figure S21) can be observed, with charged ligands as a slightly separate band towards the South East (marked by green ellipse), illustrating how PCA highlights the key sources of variation in the data. Perhaps most interesting are the ligands with different donor atoms which appear in the same area of ligand space, i.e. N,N-, O,O- and N,O-donor ligands at PC1 -5.5 - -2, PC2 -2 to 2 (Figure 3). This proximity is maintained for PC3 as well (see ESI, Figure S23) and highlights the similarity of their properties for the descriptors considered here. Indeed, while N,N-donor ligands tend to have higher metal – ligand binding energies on average than O,O-donor ligands, with N,O-ligands in-between (Table 2), the ligands in this area of chemical space are much more similar, helping to explain their placement in proximity to each other. From a structural perspective, this is perhaps not surprising, as the ligands in this area have mostly small substituents (Figure 3 top) or the donor is part of a ring/double bond (Figure 3 bottom), and the bridges are relatively short.

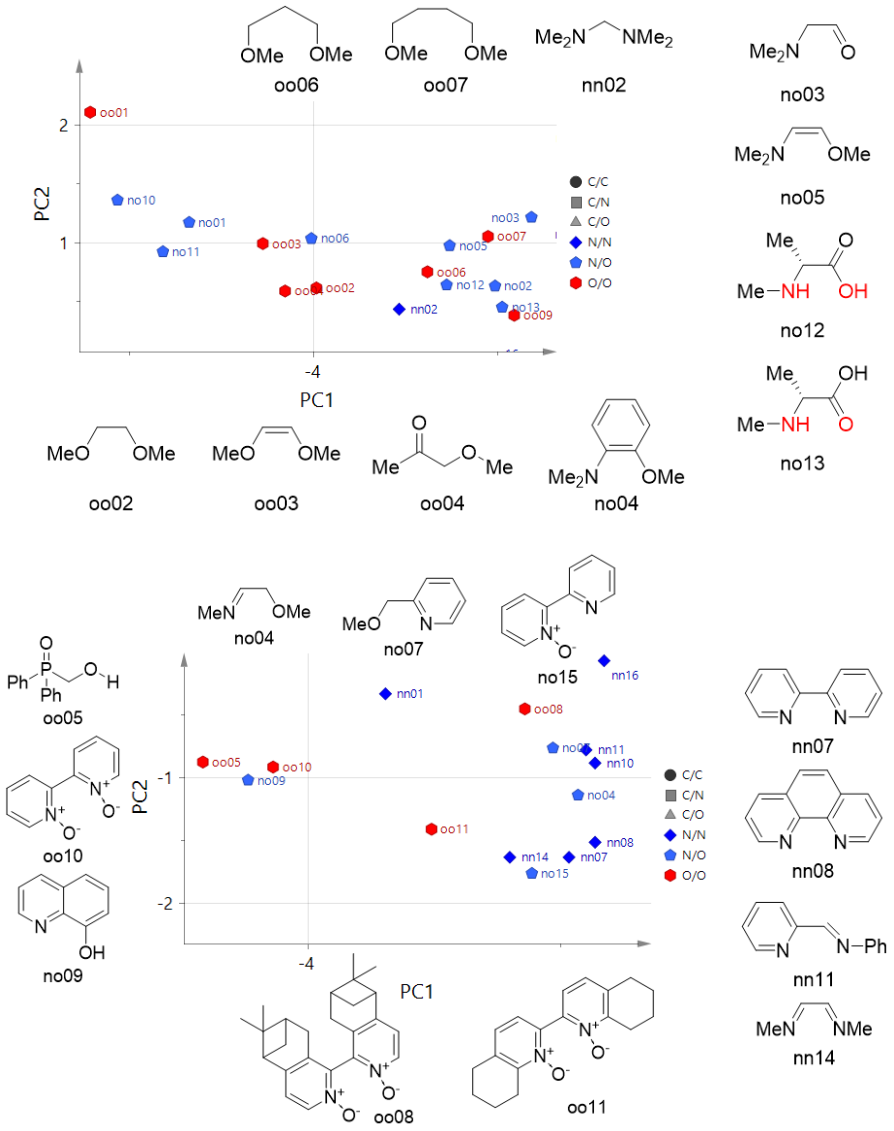


Figure 3: Focus on ligand space where N,N-, N-,O- and O,O-donor ligands overlap, including representative structures. Donors are shown in red where this is ambiguous.

While recognising and quantifying such similarities can provide useful information for experimental designs, facilitating the selection of a representative ligand set for screening, we note that the parameters used so far have treated all ligands as bidentate (or *cis*-, bis-ligating for the small number of monodentate ligands considered) and calculated binding energies have assumed full dissociation. In some cases, especially those with mixed donors, hemilability may be more likely and binding energies of individual donors to a [AuCl] fragment have been calculated to probe this further. Figure 4 shows the same PC scores as Figure 2, focussing just on C-, N- and O-donor ligands, but is coloured according to the binding energies for this fragment to individual donor atoms.

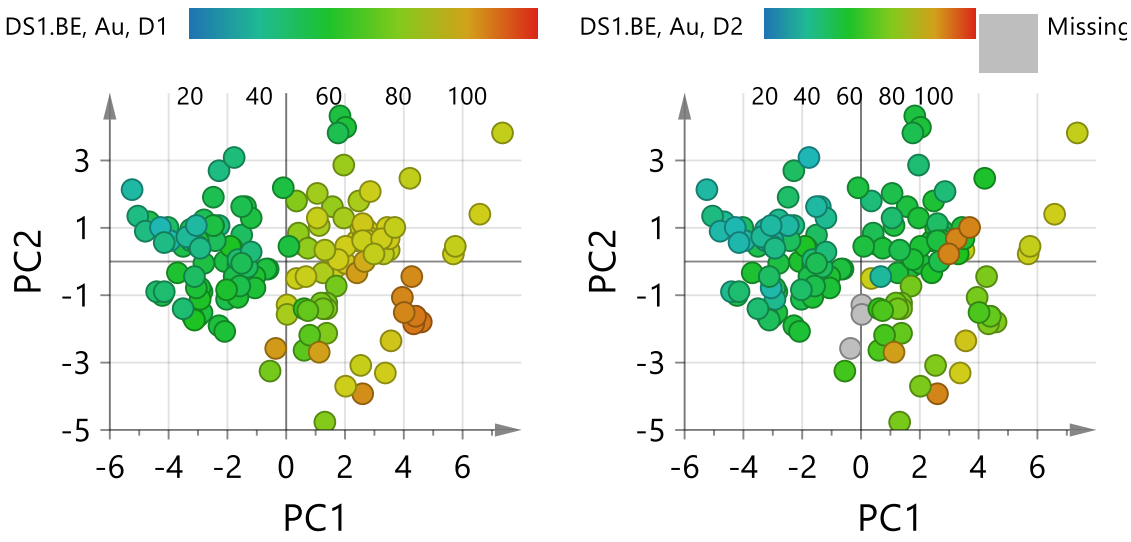


Figure 4: PC score plot, showing PCs 1 and 2, for ligands with C-, N- and O-donors, coloured by BE (Au, D) descriptors. See ESI for these plots with ligand numbers (Figure S25).

The differences between donors are most pronounced for C,N-donor ligands, predominantly found in the North East quadrant, but even for N,O-donor ligands (West), the differences are often in excess of 10 kcal mol^{-1}, suggesting a clear preference for this gold complex to bind to donor 1 (C and N respectively) over the other. This may well help to create a flexible coordination sphere, valuable in some catalytic cycles where a hemilabile ligand can "protect" a vacant site when needed.[27] Ligands with one charged donor (South East) not unexpectedly also exhibit quite large differences.

These observations align well with the mechanistic complexities for catalytic cycles such as the Ullmann coupling of aminoalcohols[4] mentioned above. This potential variability of ligand coordination will merit further consideration, including an exploration of ligand hemilability and full or partial displacement by charged substrates, especially when the latter are deprotonated in the proposed catalytic cycles. As an example, no_ch01 is a deprotonated aminoalcohol ($Me_2N(CH_2)_2O^-$), which has higher binding energies to all metal fragments than bipyridyl (nn07), phenanthroline (nn08) and acac ligands such as oo_ch02. Therefore, these compounds appear in different areas of ligand space, as noted above, but might all be present in a reaction mixture for copper-catalysed Ullmann coupling. Exploration of such competitive binding lies outside of the scope of the study reported here as it is very reaction specific whereas we pursued a general approach, but work is currently under way to combine such data-driven insights with calculated mechanistic data.

Summary and Conclusions

This mapping of bidentate ligand space with DFT-calculated steric and electronic parameters has demonstrated that ligands with different donor atoms can be combined and their properties captured within a single computational approach. Since PCA highlights the biggest sources of variation in the data, it becomes increasingly difficult to accommodate substantial variation in donors or backbones, but we have shown how further data analysis can be used to explore areas of overlap. It would also be feasible to focus on a subset of ligands, as demonstrated in earlier work on LKB-PP, for more

detailed mapping and applications to screening and prediction,[16, 17] while still retaining the same descriptors, allowing the generation of both detailed and overview maps of bidentate ligand space. The observed progression across maps according to changing donors and bridge lengths is satisfying, highlighting the importance of these two variables in ligand optimisation. Substituent effects around the donor site will also be of interest and have been explored extensively for bidentate P-donor ligands in LKB-PP_{screen},[18] but we have not investigated this variation systematically, focussing instead on representative ligands in this initial study. The parameters used have been derived from a limited number of coordination environments (square planar [$PdCl_2$] and tetrahedral [$ZnCl_2$] fragments, along with the He_8-wedges used to mimic the steric hindrance of *cis* ligands in an octahedral complex), but we expect them to have some transferability to other coordination environments. This has been demonstrated for P,P-donor ligands where the bond energy for complete dissociation of the ligands from octahedral [$Cr(CO)_4$(PP)] complexes could be predicted from a multivariate linear regression model based on these parameters.[16] Expansion to include other coordination environments is certainly feasible, but brings additional computational challenges (see ESI for a more detailed discussion).

Work is currently under way to generate experimental data, screening a range of ligands under comparable conditions, which will allow us to test the experimental utility of the present approach in the future. There are also indications that ligand hemilability and displacement by other ligands could be important, but this will require further mechanistic investigation to fully quantify, and may lead to the introduction of additional descriptors, with a focus on key reactions, in the future. Finally, we note that our focus so far has been on late transition metals and an expansion to early/harder metal centres may well enhance the insights for ligands with C-, N-, and O-donors. The work of Odom's group is of particular interest in this area.[28]

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

Generous support from the Welsh Government through the European Regional Development Fund to CatSci (NF, JDM, SNGT) is gratefully acknowledged. Interactions between this work and research projects in Leeds (CEW, BNN) and in Loughborough (AVM) were facilitated by the Dial-a-Molecule EPSRC Grand Challenge Network funded under the "Applying Statistical Methods to Chemistry Research" call (EP/K004840/1). NF also thanks the EPSRC for the award of an Advanced Research Fellowship (EP/E059376/1). AK thanks the University of Bristol and the Centre for Computational Chemistry for hosting an undergraduate research project.

Electronic Supporting Information

Full ligand labels, additional plots and diagnostic data, spreadsheets of descriptors, PC scores and loadings, sample xyz coordinates.

References

1. J. A. Gillespie, D. L. Dodds and P. C. J. Kamer, *Dalton Trans.*, 2010, **39**, 2751-2764; K. Wu and A. G. Doyle, *Nature Chem.*, 2017, **9**, 779; S. M. Mansell, *Dalton Trans.*, 2017, **46**, 15157-15174.
2. D. J. Durand and N. Fey, *Chem. Rev.*, 2019, **119**, 6561-6594.
3. T. M. Trnka and R. H. Grubbs, *Acc. Chem. Res.*, 2001, **34**, 18-29; Y. Chu, W. Heyndrickx, G. Occhipinti, V. R. Jensen and B. K. Alsberg, *J. Am. Chem. Soc.*, 2012, **134**, 8885-8895; G. Occhipinti, H. R. Bjorsvik and V. R. Jensen, *J. Am. Chem. Soc.*, 2006, **128**, 6952-6964; W. Smit, V. Koudriavtsev, G. Occhipinti, K. W. Törnroos and V. R. Jensen, *Organometallics*, 2016, **35**, 1825-1837.
4. G. O. Jones, P. Liu, K. N. Houk and S. L. Buchwald, *J. Am. Chem. Soc.*, 2010, **132**, 6205-6213; H.-Z. Yu, Y.-Y. Jiang, Y. Fu and L. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 18078-18091; G. Lefèvre, G. Franc, A. Tlili, C. Adamo, M. Taillefer, I. Ciofini and A. Jutand, *Organometallics*, 2012, **31**, 7694-7707; C. He, G. Zhang, J. Ke, H. Zhang, J. T. Miller, A. J. Kropf and A. Lei, *J. Am. Chem. Soc.*, 2013, **135**, 488-493; C. Sambiagio, S. P. Marsden, A. J. Blacker and P. C. McGowan, *Chem. Soc. Rev.*, 2014, **43**, 3525-3550; S. Sung, D. C. Braddock, A. Armstrong, C. Brennan, D. Sale, A. J. P. White and R. P. Davies, *Chem. Eur. J.*, 2015, **21**, 7179-7192; R. Giri, A. Brusoe, K. Troshin, J. Y. Wang, M. Font and J. F. Hartwig, *J. Am. Chem. Soc.*, 2018, **140**, 793-806; G. J. Sherborne, S. Adomeit, R. Menzel, J. Rabeah, A. Brückner, M. R. Fielding, C. E. Willans and B. N. Nguyen, *Chem. Sci.*, 2017, **8**, 7203-7210.
5. J. N. Harvey, F. Himo, F. Maseras and L. Perrin, *ACS Catal.*, 2019, **9**, 6803-6813.
6. J. Jover and N. Fey, *Chem. Asian J.*, 2014, **9**, 1714-1723.
7. A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**, eaau5631.
8. L. Perrin, E. Clot, O. Eisenstein, J. Loch and R. H. Crabtree, *Inorg. Chem.*, 2001, **40**, 5806-5811.
9. N. Fey, M. Garland, J. P. Hopewell, C. L. McMullin, S. Mastroianni, A. G. Orpen and P. G. Pringle, *Angew. Chem. Int. Ed.*, 2012, **51**, 118-122; N. Fey, S. Papadouli, P. G. Pringle, A. Ficks, J. T. Fleming, L. J. Higham, J. F. Wallis, D. Carmichael, N. Mézailles and C. Müller, *P, S, Si and Rel. Elem.*, 2015, **190**, 706-714; R. J. Newland, A. Smith, D. M. Smith, N. Fey, M. J. Hanton and S. M. Mansell, *Organometallics*, 2018, **37**, 1062-1073.
10. C. A. Tolman, *Chem. Rev.*, 1977, **77**, 313-348.
11. D. G. Gusev, *Organometallics*, 2009, **28**, 763-770.
12. H. Clavier and S. P. Nolan, *Chem. Commun.*, 2010, **46**, 841-861.
13. H. V. Huynh, *Chem. Rev.*, 2018.
14. N. Fey, A. Tsipis, S. E. Harris, J. N. Harvey, A. G. Orpen and R. A. Mansson, *Chem. Eur. J.*, 2006, **12**, 291-302.
15. J. Jover, N. Fey, J. N. Harvey, G. C. Lloyd-Jones, A. G. Orpen, G. J. J. Owen-Smith, P. Murray, D. R. J. Hose, R. Osborne and M. Purdie, *Organometallics*, 2010, **29**, 6245-6258.
16. N. Fey, J. N. Harvey, G. C. Lloyd-Jones, P. Murray, A. G. Orpen, R. Osborne and M. Purdie, *Organometallics*, 2008, **27**, 1372-1383.
17. J. Jover, N. Fey, J. N. Harvey, G. C. Lloyd-Jones, A. G. Orpen, G. J. J. Owen-Smith, P. Murray, D. R. J. Hose, R. Osborne and M. Purdie, *Organometallics*, 2012, **31**, 5302-5306.
18. J. Jover and N. Fey, *Dalton Trans.*, 2013, **42**, 172-181.
19. D. Livingstone, *A Practical Guide to Scientific Data Analysis*, Wiley & Sons Ltd., Chichester, UK, 2009; J. Townend, *Practical Statistics for Environmental and Biological Scientists*, John Wiley & Sons Ltd., Chichester, 2002.
20. N. Fey, *Chem. Cent. J.*, 2015, **9**, 38.
21. R. A. Mansson, A. H. Welsh, N. Fey and A. G. Orpen, *J. Chem. Inf. Model.*, 2006, **46**, 2591-2600.
22. J. D. Moseley and P. M. Murray, *J. Chem. Technol. Biotechnol.*, 2014, **89**, 623-632.
23. N. Fey, B. M. Ridgway, J. Jover, C. L. McMullin and J. N. Harvey, *Dalton Trans.*, 2011, **40**, 11184-11191.
24. A. Sengupta, A. Seitz and K. M. Merz, *J. Am. Chem. Soc.*, 2018, **140**, 15166-15169.

25. N. Fey, M. F. Haddow, J. N. Harvey, C. L. McMullin and A. G. Orpen, *Dalton Trans.*, 2009, 8183-8196.
26. R. B. Bedford, P. B. Brenner, D. Elorriaga, J. N. Harvey and J. Nunn, *Dalton Trans.*, 2016, **45**, 15811-15817; A. A. Danopoulos, T. Simler and P. Braunstein, *Chem. Rev.*, 2019, **119**, 3730-3961; K. Junge, V. Papa and M. Beller, *Chem. Eur. J.*, 2019, **25**, 122-143; Y. Y. Li, S. L. Yu, W. Y. Shen and J. X. Gao, *Acc. Chem. Res.*, 2015, **48**, 2587-2598; A. V. Malkov, D. Pernazza, M. Bell, M. Bella, A. Massa, F. Teply, P. Meghani and P. Kocovsky, *J. Org. Chem.*, 2003, **68**, 4727-4742; A. Mukherjee and D. Milstein, *ACS Catal.*, 2018, **8**, 11435-11469; G. J. Sherborne, S. Adomeit, R. Menzel, J. Rabeah, A. Bruckner, M. R. Fielding, C. E. Willans and B. N. Nguyen, *Chem. Sci.*, 2017, **8**, 7203-7210.
27. B. R. M. Lake and C. E. Willans, *Chem. Eur. J.*, 2013, **19**, 16780-16790.
28. B. S. Billow, T. J. McDaniel and A. L. Odom, *Nature Chem.*, 2017, **9**, 837-842.